64-9

ESD-TDR-64-19

Group Report 1964-4

F. C. Schweppe

Optimization of Signals

16 January 1964

Prepared under Electronic Systems Division Contract AF 19(628)-500 by

Lincoln Laboratory
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Lexington, Massachusetts

The work reported in this document was performed at Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology, with the support of the U.S. Air Force under Contract AF 19(628)-500.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

LINCOLN LABORATORY

OPTIMIZATION OF SIGNALS

F. C. SCHWEPPE

Group 28

GROUP REPORT 1964-4

16 JANUARY 1964

LEXINGTON MASSACHUSETTS

ABSTRACT

A general theory is presented for the design of optimum signal waveforms. Explicit constraints on the signal's energy, amplitude and bandwidth are possible. The development is based on optimum control theory and employs the state variable concept with differential equations as basic models. Pontryagin's Maximum Principle provides a two-point boundary value problem whose solution gives the optimum signal. A communication problem of deciding which of M signals was transmitted and a radar problem of deciding the range and range rate of a target are used as examples.

This technical documentary report is approved for distribution.

Franklin C. Hudson

Franklin C. Hudson, Deputy Chief
Air Force Lincoln Laboratory Office

1. INTRODUCTION

The design of signals for communication and radar systems is presently more of an art than a science. Many excellent signals have evolved and valuable theories are available but these are special cases which do not lead to satisfactory generalizations. Therefore to meet an obvious need, we shall introduce a methodology for signal design which encompasses a very broad class of problems, enables the incorporation of signal constraints and design criteria in a natural manner, and provides a formulation well adapted to modern computer techniques.

The signal design theory presently employed appears to be based primarily on a priori parameterizations of the signal. Optimization is then performed over the values of the parameters, (see for example, Refs. 1, 2, 3). More general results rely on optimization by inspection of special cases (see for example, Ref. 4). Signal constraints are rarely an integral part of the development and the criteria of optimality is often vague. The development of a general theory of optimum signal design has been handicapped by the analysis technique employed. Frequency domain concepts are impractical for most nonstationary problems. The Karhunen-Loeve expansion, (see Ref. 5) is a fine theoretical tool but often obscures the real issues by its generality. In addition, integral equations are often obtained and they are usually not well suited to computer investigation. Several years ago, control theory was in this same situation as it relied on frequency domain techniques, optimization of parameterized systems and a wide variety of special purpose approaches. However, a return to the time domain and the state variable concepts of classical physics, and an extension of the calculus of variations called Pontryagin's Maximum Principle, provided a basis for a general approach which is presently called optimal control theory. Non-stationary stochastic processes became a minor extension when interest was confined to finite order Markov processes, a model of sufficient generality for most problems. The resulting differential and difference equations are well adapted to computer investigation.

This report merely marries the concepts of optimal control theory to the design of signals for communication and radar systems. The only addition required is the incorporation of statistical decision theory in an appropriate form. We do not attempt full consummation of the wedding, but rather, use examples to indicate possible progeny.

The present theory is an outgrowth of studies on the design of optimum amplitude and frequency modulations for a radar. The amplitude modulation work was done with Donald Gray.* These results are presently being documented and will be available shortly as explicit examples of how the theory can be employed in realistic problems. Dr. Robert Price† introduced the author to communication applications, motivated the present writing and provided many valuable comments.

Scalars are denoted by lower case letters, with or without subscripts. Vectors are underlined lower case letters, with or without subscripts. Matrices are denoted by capital letters. The exceptions to these rules are so indicated in the text. All vectors are column vectors. Transpose is indicated by a prime. "$| \ |$" denotes the determinant and "tr" the trace of a matrix. Henceforth, the term waveform, applies to the signal to be designed and x always refers to this waveform. The quantities observed by the receiver are always denoted by z. Covariance matrices are denoted by Σ and variances by σ^2 with appropriate subscripts. We use the same symbol for both a random variable and a sample of the random variable and rely on the text to furnish the necessary distinction. Discrete white noise is a sequence of zero mean, stationary independent Gaussian random variables. Continuous time white noise stochastic processes are zero mean, stationary Gaussian processes with uniform spectrum and are employed in a nonrigorous manner.

---

* MIT, Lincoln Laboratory (Summer Staff).

† MIT, Lincoln Laboratory.

2. TWO EXAMPLES

A simple communication problem and a simple radar problem are stated. These two examples will be carried through the general development to provide a vehicle for illustrating the methodology of waveform design and indicating its generality.

Consider the following communication problem. A receiver knows that one of M waveforms

$$x_k(t), \; 0 \le t \le T, \qquad k = 1, \ldots, M \tag{2.1}$$

was transmitted. The receiver observes the transmitted waveform corrupted by additive noise and must decide which of the M waveforms was transmitted. Let $z(t)$ be the observed signal. Then

$$z(t) = \sum_{k=1}^{M} \alpha_k x_k(t) + \eta(t) \tag{2.2}$$

where $\eta(t)$ is the observation noise and

$$\alpha_k = 0 \qquad k \neq k_0$$

$$\alpha_k = 1 \qquad k = k_0$$

where $x_{k_0}(t)$ is the waveform actually transmitted. Assume $\eta(t)$ is white noise. The choice of possible $x_k(t)$ is restricted by constraints on total energy and bandwidth. It is desired to choose the set of functions, $x_k(t)$, $k = 1, \ldots, M$ which provide the best probability of making the correct decision as to which $\alpha_k = 1$.

Consider the following radar problem. A target is moving with respect to a radar at a range, $\mathcal{r}(t)$, given by

$$\mathcal{r}(t) = \alpha_1 + \alpha_2 t \tag{2.3}$$

where α_1 and α_2 are the range and range rate at time zero.* Assume the transmitted radar signal is a sequence of many short pulses of amplitude $x(n)$ where n indexes time and that the radar receiver estimates only the range, $r(n)$, at time n from each pulse. Let $\tilde{\eta}(n)$ denote the error made in this estimate of range. A model for the quantity, $\tilde{z}(n)$, observed from each pulse is then

$$\tilde{z}(n) = r(n) + \tilde{\eta}(n)$$

where

$$E(\tilde{\eta}^2(n)) = \sigma^2/x^2(n)$$

where σ^2 is determined by the background noise and the bandwidth of the transmitted pulse. Since $x(n)\,\tilde{z}(n)$ contains the same information as $\tilde{z}(n)$, the receiver's observations can be modeled as

$$z(n) = x(n)\,r(n) + \eta(n)$$

where

$$E(\eta^2(n)) = \sigma^2 .$$

Further, assume $\eta(n)$ is discrete white noise. For convenience, we will use a continuous time approximation to this model; that is,

$$z(t) = x(t)\,r(t) + \eta(t) \tag{2.4}$$

where $\eta(t)$ is white noise with power, σ^2.† The choice of possible $x(t)$ is restricted by constraints on its peak amplitude and total energy. It is desired to determine the $x(t)$ which provides the best estimate of α_1 and α_2.

---

* A $r(t)$ of the form $\sum_{k=1}^{M} \alpha_k \phi_k(t)$ is a trivial extension.

† Under appropriate assumptions on bandwidth and ability to resolve phase, the model of Eq. (2.4) is equivalent to that where $x(t)$ is the amplitude modulation of a high frequency carrier and $z(t)$ is the (complex) r-f signal. However, the proof of this equivalence is lengthy and is omitted.

3. MEAN INFORMATION CONTENT OF RECEIVED SIGNAL

Communication and radar systems receive transmitted signals which have passed through a channel which may change the signal's form and/or introduce additive observation noise. On the basis of the receiver's observations, a decision of some nature is desired. The prime difference between communication and radar systems is the nature of this decision. Communication systems must decide which signal was transmitted. Radar systems must decide the nature of the channel; for example, the distance between the radar and the target which reflects the signal. In both cases classical statistical decision theory provides a basis for making and evaluating these decisions. The optimum transmitted signal is the signal which results in the "fewest" decision errors.

The probability distribution of the received signal obviously specifies the decision errors. However, for our development we employ only a particular function of the probability distributions rather that the distributions themselves. This function is the mean information contained in the observed signal. Our discussions are based on Ref. 6 although we use neither the same generality nor degree of rigor.

Consider the observable random variable, z. Assume we know that z has one of two probability density functions, $f_1(z)$ or $f_2(z)$. On the basis of observing a particular realization of z, we want to decide between the two hypotheses, H_j, $j = 1, 2$ where H_j is the hypothesis that $f_j(z)$ is the probability density of z. Symbolically,

$$H_1 \quad : \quad f_1(z)$$

$$H_2 \quad : \quad f_2(z) \, .$$

Let $I(1{:}2/z)$ denote the mean information contained in an observation, z, for discrimination in favor of H_1 against H_2. This mean information is defined as the expected value, assuming H_1 is true, of the logarithm of the likelihood ratio.

$$I(1{:}2/z) = \int f_1(z) \log \frac{f_1(z)}{f_2(z)} \, dz \qquad (3.1)$$

A few comments on this definition of mean information are provided in the appendix. For simplicity we call, $I(1:2/z)$ the information rather than mean information.

If we make vector observations, $\underline{z}' = [z_1, \ldots, z_N]$,

$$I(1:2/\underline{z})$$

$$= \int\int \cdots \int f_1(\underline{z}) \log \frac{f_1(\underline{z})}{f_2(\underline{z})} \, dz_1 \cdots dz_N \tag{3.2}$$

or more concisely,

$$I(1:2/\underline{z}) = \int f_1(\underline{z}) \log \frac{f_1(\underline{z})}{f_2(\underline{z})} \, d\underline{z} \tag{3.3}$$

Continuous time observations, $z(t)$, $0 \leq t \leq T$, are a similar extension. Since discrete and continuous observations involve basically the same concepts, we abuse notation and let the symbol $I(1:2/\underline{z})$ also apply to the continuous time case. Information is additive in the following sense. If $z_1, \ldots, z_N$ are independent samples from the same distribution, then

$$I(1:2/\underline{z}) = \sum_{n=1}^{N} I(1:2/z_n) \quad . \tag{3.4}$$

Now consider the case where the two distributions, $f_1(\underline{z})$ and $f_2(\underline{z})$ differ only with respect to the values of a vector parameter, $\underline{\theta}$. That is,

$$H_1 \quad : \quad f(\underline{z}/\underline{\theta}_1)$$

$$H_2 \quad : \quad f(\underline{z}/\underline{\theta}_2) \; .$$

If $\underline{\theta}_1 = \hat{\underline{\theta}}$ and $\underline{\theta}_2 = \hat{\underline{\theta}} + \Delta\underline{\theta}$, where $\Delta\underline{\theta}$ is small, then under sufficient regularity conditions,

$$2\, I(1:2/\underline{z}) = \Delta\underline{\theta}' \mathcal{I} \Delta\underline{\theta} \tag{3.5}$$

where prime denotes transpose and $\mathcal{I}$ is the positive definite, Fisher information matrix whose elements, ι_{jk}, are given by

$$\iota_{jk} = \int \frac{1}{f(\underline{z},\hat{\underline{\theta}})} \frac{\partial f(\underline{z},\hat{\underline{\theta}})}{\partial \theta_j} \frac{\partial f(\underline{z},\hat{\underline{\theta}})}{\partial \theta_k} dz \quad . \tag{3.6}$$

See Sec. 2.6 of Ref. 6 for derivations of Eq. (3.5) and Eq. (3.6).

If z(t) is of the form

$$z(t) = v(t,\underline{\theta}) + \xi(t) \qquad 0 \leq t \leq T \tag{3.7}$$

where $\xi(t)$ is white noise with power σ^2, then it is not difficult to show

$$\mathcal{I} = \frac{1}{\sigma^2} \int_0^T \underline{\phi}(t)\underline{\phi}'(t)dt \tag{3.8}$$

where the elements, $\phi_k(t)$, of $\underline{\phi}(t)$ are

$$\phi_k(t) = \frac{\partial v(t,\underline{\theta})}{\partial \theta_k} \tag{3.9}$$

If $\mathcal{I}(t)$ denotes the Fisher information matrix for $z(\tau)$, $0 \leq \tau \leq t$, then

$$\frac{d\mathcal{I}(t)}{dt} = \frac{1}{\sigma^2} \underline{\phi}(t)\underline{\phi}'(t) \tag{3.10}$$

where the left hand side denotes the matrix of the time derivatives of the elements of $\mathcal{I}(t)$. If $\Sigma_{\underline{\theta}}$ denotes the covariance matrix of any unbiased estimate of $\underline{\theta}$, then

$$\Sigma_{\underline{\theta}} \geq \mathcal{I}^{-1} \tag{3.11}$$

where the inequality means the difference is a positive definite matrix.* If Eq. (3.7) is linear; that is,

$$v(t, \underline{\theta}) = \sum \theta_k \phi_k(t) \tag{3.12}$$

then equality in Eq. (3.11) is obtained when a minimum variance, unbiased estimate is used.

The model, Eq. (3.7) and its implications Eqs. (3.10) and (3.11), are well-known. However, the following extension enables the modeling of a far wider range of problems. For simplicity, we consider only linear systems. Assume vector observations, $\underline{z}(t)$ of the form

$$\underline{z}(t) = C(t)\underline{y}(t) + \underline{\xi}_2(t) \tag{3.13}$$

$$\frac{d}{dt}\underline{y}(t) = A(t)\underline{y}(t) + B(t)\underline{\xi}_1(t) \tag{3.14}$$

where $\underline{y}(t)$ is a vector, $A(t), B(t)$, and $C(t)$ are known, possibly time-varying matrices, and $\underline{\xi}_1(t)$ and $\underline{\xi}_2(t)$ are vector white noise processes, independent of each other with

$$E[\underline{\xi}_1(t)\underline{\xi}_1'(t)] = \Sigma_{\underline{\xi}_1} \tag{3.15}$$

$$E[\underline{\xi}_2(t)\underline{\xi}_2'(t)] = \Sigma_{\underline{\xi}_2} \tag{3.16}$$

Equations (3.13) and (3.14) are the state variable representation for finite dimensional, time-varying, linear processes.† y(t) is a Markov process. By appropriate choice of driving noise, $\underline{\xi}_1(t)$, and system dynamics, $A(t), B(t), C(t)$, a wide variety of

* This inequality is often called either the Cramer-Rao or Information Inequality and is used as a basis for defining the efficiencies of estimates, (see Sec. 3.5 of Ref. 6 or Sec. 12.6 of Ref. 7). Reference 6 also extends the inequality to the case of biased estimates.

† Reference 8 discusses the state variable concept in great detail.

deterministic and stochastic signals and correlated observation noise can be modeled. Equation (3.14) is a linear system driven by white noise and is an extension of the "pre-whitening" filter concept discussed in Ref. 9. If $\mathcal{I}(t)$ denotes the Fisher information matrix for the vector observations $\underline{z}(\tau)$, $0 \leq \tau \leq t$ of Eq. (3.13), it can be shown that

$$\frac{d\mathcal{I}(t)}{dt} = -\mathcal{I}(t)A(t) - A'(t)\mathcal{I}(t) - \mathcal{I}(t)B(t)\Sigma_{\underline{\xi}_1} B'(t)\mathcal{I}(t) + C'(t)\,\Sigma^{-1}_{\underline{\xi}_2} C(t) \tag{3.17}$$

The $\mathcal{I}(t)$ of Eq. (3.17) is the information available on y(t) at time t. This is related to Eqs. (3.7) and (3.10) by associating $\underline{\theta}$ with $\underline{y}(t)$. References 10 and 11 made the original investigations of equations such as Eq. (3.17). Reference 12 uses a different method of derivation.* The initial conditions for Eq. (3.17); i.e., $\mathcal{I}(0)$, depend on the information available before the receipt of the signal. Equation (3.17) is a matrix Ricatti equation. The first two terms show how the information available on $\underline{y}(t)$ varies with time (with $\underline{y}(t)$) when no driving noise is present ($\Sigma_{\underline{\xi}_1} = 0$) and when no observations are made($\Sigma^{-1}_{\underline{\xi}_2} = 0$); that is, how $\mathcal{I}(0)$ evolves. The third term indicates the rate of decrease in information caused by the driving noise while the fourth term gives the rate of increase of information from the observations.

Now consider the basic problem of parameter estimation; that is, deciding on an explicit estimate for the values of some parameter set. Let $\hat{\underline{\theta}}$ denote the estimate of $\underline{\theta}$ made from the observations $\underline{z}$. For a high signal-to-noise ratio,† $\Delta\underline{\theta} = \hat{\underline{\theta}} - \underline{\theta}$, is small and the Fisher information matrix, $\mathcal{I}$, measures the information contained in $\underline{z}$ for discriminating between $\underline{\theta}$ and $\hat{\underline{\theta}}$.

---

* These references actually derive the equation for the covariance matrix $\Sigma_y(t)$ of the minimum variance (unbiased) estimate of $\underline{y}(t)$. However, since the model is linear, $\mathcal{I}(t) = \Sigma_y^{-1}(t)$.

† The necessary magnitude of the signal-to-noise ratio is intimately related to the dynamics of the signal. The ideas of Ref. 13 are valuable in considering such problems.

The radar problem of Sec. 2 is one of parameter estimation and the Fisher information is thus appropriate.* For the example continued through the following sections, the observation noise, $\eta(t)$, of Eq. (2.4) is assumed white. Thus, from Eq. (3.10) (or a very special case of Eq. (3.17),

$$\frac{d\mathcal{I}(t)}{dt} = \frac{x^2(t)}{\sigma^2}\begin{bmatrix} 1 & t \\ t & t^2 \end{bmatrix} \tag{3.18}$$

However, to illustrate the wide range of models possible using Eqs. (3.13) and (3.14), temporarily assume that

$$\eta(t) = \eta_1(t) + \eta_2(t) + \eta_3 \tag{3.19}$$

where $\eta_1(t)$ is a second-order, nonstationary Markov process of the form ,

$$\frac{d^2\eta_1(t)}{dt^2} + a_1(t)\frac{d}{dt}\eta_1(t) + a_2(t)\eta_1(t) = \text{white noise}$$

---

* Since the model is a linear, Gaussian process, an assumption of high signal-to-noise is not needed as $\mathcal{I}^{-1}$ gives the variance with which the parameters can be estimated.

$\eta_2(t)$ is white noise and η_3 is an unknown constant representing an unknown mean value. Define

$$\frac{d}{dt}\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -a_2(t) & -a_1(t) \end{bmatrix}\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \xi_1(t)$$

$$\frac{d}{dt} y_3(t) = 0$$

where $\xi_1(t)$ is white noise and where we associate $\eta_1(t)$ with $y_1(t)$, $\eta_2(t)$ with $\xi_2(t)$ (white noise) and η_3 with $y_3(t)$. We can combine this observation noise with the radar example by letting the state variables $y_4(t), \ldots, y_5(t)$ correspond to the constants, α_1 and α_2 of Eq. (2.4). Thus

$$\frac{dy_j(t)}{dt} = 0 \qquad j = 4, 5$$

and in terms of Eq. (3.13) and (3.14), the radar example with the noise of Eq. (3.19) is

$$A(t) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -a_2(t) & -a_1(t) & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$B(t) = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$C(t) = \begin{bmatrix} 1 \\ 0 \\ 1 \\ x(t) \\ tx(t) \end{bmatrix}$$

However, instead of representing the range as in Eq. (2.3), we could define,

$$y_4(t) = r(t)$$
$$y_5(t) = \frac{d}{dt} r(t)$$

obtain the differential equation,

$$\frac{d}{dt}\begin{bmatrix} y_4(t) \\ y_5(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} y_4(t) \\ y_5(t) \end{bmatrix}$$

and then consider the problem of estimating the range, $y_4(t)$, and range rate, $y_5(t)$, at time t rather than α_1 and α_2, the range and range rate at time zero. With this approach, B(t) is still unchanged while

$$A(t) = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ -a_2(t) & -a_1(t) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

and

$$C(t) = \begin{bmatrix} 1 \\ 0 \\ 1 \\ x(t) \\ x(t) \end{bmatrix}$$

With this second formulation, extension to targets with stochastic motion can be handled by changing B(t) and $\underline{\xi}_1(t)$. Reference 14 considers the use and implications of such models in filter design.

The communication type problem requires a decision which differs from parameter estimation as we want to decide which, if any, signal was transmitted. This is not a choice between a parameter, $\underline{\theta}$, and a small perturbation, $\underline{\theta} + \Delta\underline{\theta}$. Thus the Fisher information matrix does not necessarily apply and we must return to the basic definition of information. However, a result analogous to the Fisher information matrix is obtained.

For the communication problem of Sec. 2 the observation noise, $\eta(t)$, of Eq. (2.2) is white. We begin with a discrete time version of Eq. (2.2)

$$z(n) = \phi(n) + \eta(n) \qquad n = 1, \ldots, N \tag{3.20}$$

where

$$\phi(n) = \sum_{k=1}^{M} \alpha_k x_k(n) \tag{3.21}$$

and $\eta(n)$ is discrete white noise with variance, σ^2. Let $\underline{z}$, denote the vector of the z(n). Then

$$f(\underline{z}) = \frac{1}{(2\pi\sigma^{2N})^{\frac{1}{2}}} \exp\left\{ \frac{\sum_{n=1}^{N} (z(n) - \phi(n))^2}{-2\sigma^2} \right\} .$$

We want to decide which, if any, of the α_k is one. M hypothesis tests of the form

$$\begin{aligned} H_1 &: \alpha_k = 0 \qquad \text{all } k \\ H_2 &: \alpha_k = 0 \begin{cases} 0 & k \neq j \\ 1 & k = j \end{cases} \end{aligned} \tag{3.22}$$

for j = 1, . . . , M would solve the problem (provided H_2 was accepted only once). Let $I_j(1:2/\underline{z})$ denote the information for the j^{th} of these tests. Using Eq. (3.2), it is

not difficult to show that

$$2I_j(1:2/\underline{z}) = \frac{1}{\sigma^2}\sum_{n=1}^{N} x_j^2(n) . \tag{3.23}$$

Thus for this case, $I_j(1:2/\underline{z})$ is proportional to the waveform's energy. Equation (3.23) is the information which controls type 1 errors; that is, deciding that $\alpha_j = 0$ when in fact, $x_j(t)$ was transmitted. The type 2 error of deciding $\alpha_j = 1$ when in fact $x_k(t)$ was transmitted is also important and is controlled by the information with respect to hypotheses of the form

$$\begin{aligned} H_1 \quad &: \quad \alpha_i \begin{cases} = 0 & i \neq j \\ = 1 & i = j \end{cases} \\ H_2 \quad &: \quad \alpha_i \begin{cases} = 0 & i \neq k \\ = 1 & i = k \end{cases} \end{aligned} \tag{3.24}$$

for $j, k = 1, \ldots, M,\ j \neq k$. Let $I_{jk}(1:2/\underline{z})$ denote the information for one of these tests. Then

$$2I_{jk}(1:2/\underline{z}) = \frac{1}{\sigma^2}\sum_{n=1}^{N} (x_j(n) - x_k(n))^2 \tag{3.25}$$

Thus for this case, $I_{jk}(1:2/\underline{z})$ is proportional to the "distance" between the waveforms. Now define the M x M matrix, $\hat{I}$ with elements $\hat{\imath}_{jk}$ by

$$\hat{I} = \frac{1}{\sigma^2}\sum_{n=1}^{N} \underline{x}(n)\underline{x}'(n) \tag{3.26}$$

where the $x_j(n)$ are the elements of $\underline{x}(n)$. By inspection

$$2 I_{jk}(1{:}2/\underline{z}) = \hat{\iota}_{jj} - 2\hat{\iota}_{jk} + \hat{\iota}_{kk}$$

$$2 I_{j}(1{:}2/\underline{z}) = \hat{\iota}_{jj}\ .$$

Thus $\hat{I}$ contains all the terms that control the probabilities of the type 1 and 2 errors. Extending Eq. (3.26) to the continuous time case gives

$$\hat{I} = \frac{1}{\sigma^2} \int_0^T \underline{x}(t)\underline{x}'(t)\,dt \qquad (3.27)$$

which has exactly the same form as Eq. (3.10). Thus $\hat{I}$ of Eq. (3.27) is the same as the Fisher information matrix for estimating the α_k of Eq. (2.2). A finite $\hat{I}(0)$ implies an a priori Gaussian distribution on the transmitted signals or the incorporation of the first two moments of such an a priori distribution.

Since our communication example involves a linear Gaussian process, the above result is almost a foregone conclusion. We outlined the basic arguments as they also apply in the general case. To decide which of M signals was transmitted, we evaluate the information on M tests such as Eq. (3.22) and $\frac{M(M-1)}{2}$ tests such as Eq. (3.24). All these quantities are then summarized into a single matrix equation, $\hat{I}$, of which Eq. (3.27) is an example. In general, $\hat{I}$, need not be the Fisher information matrix. However, it is easy to see that using Markov rather than white observation noise in the communication example does not affect the relation between $\hat{I}$ and the Fisher matrix. The incorporation of observation noise such as Eq. (3.19) into the communication problem is therefore a straightforward extension of the radar example.

In future discussions we use the symbol I for either the Fisher information matrix or $\hat{I}$ and simply call I, the information matrix.

The main theme of this section is the illustration of how the information content of an observed signal can be represented in state variable notation. For our examples, the elements of the information matrix $I(t)$ are the state variables and they are defined by differential (or difference) equations such as Eq. (3.10) or (3.17). We have employed the information measure of Eq. (3.1), because it seems to be the most useful

concept now available. However, this definition is not basic to our general theory of waveform design. The crux of the problem is the representation, in a state variable formulation, of the information content of the received signal. Any appropriate definition of information can be used.

The radar and communication examples of Sec. 2 are developed still further in the following sections. However, even though these two examples do illustrate many facets of the theory, they are both limited to the case of observing a deterministic signal in the presence of additive Gaussian noise. To show that this is not an inherent assumption, the information content of a signal is now evaluated for a simplified version of a laser radar.* This example, however, is not continued thru the rest of the report and can be by-passed if desired.

Consider the following variation of the laser radar model discussed in Ref. 15. The radar transmitter is a photon source whose average power output is modulated by the waveform $x(t)$. The receiver counts the number of photons which arrive. Let $z(n\Delta t)$ be the number of photon arrivals during the n^{th} time interval of length Δt. $z(n\Delta t)$ is assumed to be a sample of a Poisson random variable, where

$$f(z) = \frac{(\lambda(n\Delta t))^z e^{-\lambda(n\Delta t)}}{z!} \tag{3.28}$$

where $\lambda(n\Delta t)$ is the average rate of photon arrivals during the time interval Δt. The model for $\lambda(n\Delta t)$ is

$$\lambda(n\Delta t) = \lambda_o + \beta x(n-\tau(n)) \tag{3.29}$$

where β is the magnitude reflected signal, λ_o is the mean rate of the background radiation (noise) and where $\tau(n)$ is the time delay between transmission and receipt of the signal. As in the radar example of Sec. 2, we assume

$$\tau(n) = \alpha_1 + \alpha_2 n \tag{3.30}$$

* We have also investigated stochastic signals such as arise from multipath, (see, Ref. 16 or Chapt. 11 of Ref. 17). However, this topic is deferred to a later report.

where α_1 and α_2 are related to the target's range and range rate. We have also made the assumption that Δt is small enough the $x(t-\tau(t))$ is essentially constant over a time span Δt. We want to estimate α_1 and α_2 (assuming λ_o and β are known) and therefore use the Fisher information matrix as the measure of information content. From Eq. (3.28)

$$\frac{\partial f(z)}{\partial \alpha_j} = f(z) \left(\frac{z(n\Delta t)}{\lambda(n\Delta t)} - 1 \right) \frac{\partial \lambda(n\Delta t)}{\partial \alpha_j} \quad . \tag{3.31}$$

After manipulation, Eq. (3.6) becomes

$$i_{jk}(n\Delta t) = \lambda(n\Delta t) \frac{\partial \lambda(n\Delta t)}{\partial \alpha_j} \quad \frac{\partial \lambda(n\Delta t)}{\partial \alpha_k} \tag{3.32}$$

where the $i_{jk}(n\Delta t)$ are the elements of $I(n\Delta t)$, the information matrix for $z(n\Delta t)$. Now assume α_1 and α_2 are small. Then

$$\frac{\partial \lambda(n\Delta t)}{\partial \alpha_1} = \beta \frac{dx(n)}{dt}$$

and

$$\frac{\partial \lambda(n\Delta t)}{\partial \alpha_2} = \beta n \frac{dx(n)}{dt}$$

where $dx(n)/dt$ is $dx(t)/dt$ evaluated at $t = n$. Thus

$$I(n\Delta t) = (\lambda_o + \beta\, x(n)) \frac{\beta\, dx(n)}{dt} \begin{bmatrix} 1 & n \\ n & n^2 \end{bmatrix} \quad . \tag{3.33}$$

Equation (3.33) is the information obtained during one time interval. If we assume independence between time intervals, the total information is the sum of the $I(n\Delta t)$. Thus, Eq. (3.33) is a discrete time analog for $dI(t)/dt$. Note the similarity between Eq. (3.33) and Eq. (3.18). The assumption of low signal-to-noise ratio, $\lambda_o >> \beta\, x(n)$ as made in Ref. 15, obviously simplifies the problem.

4. CRITERIA

Since information has been defined by a positive definite matrix, a maximum matrix occurs if the difference between the maximum matrix and any other attainable matrix is positive definite. Unfortunately, such a maximum is rarely possible and it is usually necessary to choose a scalar criterion. The variety of conflicting criteria often leaves the designer impaled on the horns of a dilemma but this is part of the price one must pay when asking for a mathematically optimum waveform. The concept of noninferior controls (waveforms) as discussed in Ref. 18, has proven valuable in these situations, but here we merely illustrate the wide range of possible criterion and discuss how they are handled.

Consider parameter estimation. In general, the waveform that gives the best estimate of one parameter, say range, does not correspond to the best waveform for estimating a different parameter, say range rate. If one parameter is most important, a reasonable criterion is to maximize the information, $\mathcal{I}_1$, on this one parameter. For linear Gaussian processes such as the radar example of Sec. 2,

$$\sigma_{11}^2 = \frac{1}{\mathcal{I}_1}$$

where σ_{11}^2 is the variance of the estimate of the parameter.* A useful extension of this idea is to maximize the information on one parameter under a constraint that the information on another is equal to or greater than some specified level. Maximization of information on some weighted combination of the parameters is also reasonable. The determinant and trace are useful measures of the magnitude of a positive definite matrix which have the desirable property of being invariant with respect to linear transformations of the parameters. For linear problems, the trace of $\mathcal{I}^{-1}$ is the expected value of the square of the magnitude of the error vector while the determinant

* Maximization of the information on, say, the range parameter, makes the second partial derivative with respect to range of the magnitude of the ambiguity function most negative and thereby locally "peaks" the ambiguity function in the range dimension, (see Ref. 4).

of $\mathcal{I}^{-1}$, $|\mathcal{I}^{-1}| = 1/|\mathcal{I}|$ is a measure of the volume of the error ellipsoid. In radar problems, the motion of the target is often such that the determinant has the useful property of being independent of the time at which the state of the target is estimated. Equation (2.3) is an example of such motion.

A "good" information matrix for the communication type problem is one with large main diagonal elements (Eq. (3.23) large) and small off-diagonal elements (Eq. (3.25) large). If the time duration, T, of the signal is fixed, a natural choice of criterion is to require

$$\iota_{kk}(T) \geq \text{constant} \qquad k = 1, \ldots, M \tag{4.1}$$

and then maximize the minimum

$$\iota_{kk}(T) - 2\iota_{kj}(T) + \iota_{jj}(T) \qquad j, k = 1, \ldots, M \quad j \neq k \tag{4.2}$$

Of course, other criteria abound. For example, the determinant or trace or a weighted combination of the $\iota_{kk}(T)$ and $\iota_{jk}(T)$ raised to some power could be used. However, such approaches do not appear to have a natural interpretation.

In many problems we have a specified performance which must be achieved. For example, the parameter estimate variance or the type 1 and type 2 error probabilities may be required to be equal to or below certain specified levels. We might then want the signal of minimum time duration, T, which satisfies the specifications. The criterion is then to minimize T. Instead of minimizing T, we might want to minimize the signal energy or its bandwidth (see the next section).

Often, only a portion of the information matrix is germane to the choice of criteria. For example, in Sec. 3 we discussed how extra state variables (nuisance variables) are introduced to handle nonwhite observation noise. In such cases we partition the information matrix

$$\mathcal{I} = \begin{bmatrix} \mathcal{I}_{11} & \mathcal{I}_{12} \\ \mathcal{I}_{21} & \mathcal{I}_{22} \end{bmatrix}$$

corresponding to a partitioning of the state vector, $\underline{y}$, into sub vectors $\underline{y}_1$ and $\underline{y}_2$ where the $\underline{y}_2$ are nuisance variables. Let I_1 denote the information available on just $\underline{y}_1$. Then by the matrix partitioning theorems or by the theory of testing a partitioned hypothesis, (see Ref. 6),

$$I_1 = I_{11} - I_{12} I_{22}^{-1} I_{21}$$

I_1 is the information matrix to be used in our criteria.

All of the above criteria can be formulated in state variable notation . Define a state variable, $\ell(t)$, by a differential equation of the form,

$$\frac{d\ell(t)}{dt} = g(I(t), t) \tag{4.3}$$

It is desired to maximize $\ell(T)$ under various constraints on $I(T)$. The following examples illustrate the techniques. To maximize, $|I(T)|$ we might use

$$\ell(T) = \log |I(T)|$$

and

$$\frac{d\ell(t)}{dt} = \operatorname{tr} I^{-1}(t) \frac{dI(T)}{dt} \tag{4.4}$$

For the criterion of Eqs. (4.1) and (4.2) we invoke Eq. (4.1) and set

$$\iota_{kk}(T) - 2\iota_{kj}(T) + \iota_{jj}(T) \geq \ell(T) \tag{4.5}$$

and

$$\frac{d\ell(t)}{dt} = 0 \ . \tag{4.6}$$

To minimize the time duration T, we set

$$\frac{d\ell(t)}{dt} = -1$$

and apply the desired constraints on $I(T)$.

5. SIGNAL CONSTRAINTS

All signals are limited in time duration, bandwidth occupied, peak amplitude and total energy. However, it is often necessary to employ only the critical constraints in the mathematical formulation. For example, frequency modulation affects only the bandwidth of the transmitted signal. A total energy or peak amplitude constraint may result in an infinite bandwidth signal but the bandwidth of the actual system may be large enough to validate the use of the result.

It is useful to introduce a function, u(t), called the control function. This function specifies the waveform through either of the two relationships

$$x(t) = u(t) \tag{5.1}$$

or

$$x(t) = w_1(t) \tag{5.2}$$

where $w_1(t)$ is one element of the state vector $\underline{w}(t)$ defined by the vector equation,

$$\frac{d}{dt}\underline{w}(t) = D_1\underline{w}(t) + D_2u(t) \tag{5.3}$$

Two basic forms of constraints are discussed, a strict inequality constraint on the control function,

$$\epsilon_1 \le u(t) \le \epsilon_2 \tag{5.4}$$

and integral constraints of the form*

$$\int_0^T s_k(\underline{w}(t), u(t), t)dt \le \epsilon_k \tag{5.5}$$

Now consider how Eqs. (5.1), thru (5.5) are used to express physical constraints. A total energy constraint on the waveform is obtained from

---

* ϵ is used as a generic constant which is rarely related from one equation to the next.

$$x(t) = u(t)$$

$$\int_0^T u^2(t)dt \le \epsilon$$

A peak amplitude constraint on the waveform is obtained from

$$x(t) = u(t)$$

$$\epsilon_1 \le u(t) \le \epsilon_2 \quad .$$

Let $X(\omega)$ be the Fourier transform of $x(t)$. The second moment of $X(\omega)$ is a measure of bandwidth. By Parseval's theorem

$$\int_{-\infty}^{\infty} \omega^2 X^2(\omega)d\omega = \int_0^T \left(\frac{dx(t)}{dt} \right)^2 dt$$

to within a constant of proportionality. Therefore, a second moment bandwidth constraint on the waveform is obtained from

$$x(t) = w(t)$$

$$\frac{d}{dt} w(t) = u(t)$$

$$\int_0^T u^2 dt \le \epsilon \quad .$$

By using general cases of Eq. (5.2) and (5.3) and integral square constraints on both $u(t)$ and the elements of $\underline{w}(t)$, a wide range of bandwidth constraints of this type can be developed. If $x(t)$ is a frequency modulation with a large modulation index, then the significant side bands of the transmitted signal are contained within the range of values spanned by $x(t)$, (see Ref. 19). Thus,

$$x(t) = u(t)$$

$$\epsilon_1 \leq u(t) \leq \epsilon_2$$

could also be a bandwidth constraint.

The preceding constraints encompass a wide variety of practical problems. However, they do not include problems such as the combination of peak amplitude and bandwidth constraints illustrated by

$$w(t) = x(t)$$

$$\frac{dw(t)}{dt} = u(t)$$

$$\int_0^T u^2 dt \leq \epsilon$$

$$\epsilon_1 \leq w(t) \leq \epsilon_2 \quad .$$

The last expression is called a phase coordinate constraint. Such constraints can also be handled, albeit with more difficulty (see the next section).

As with the criteria of Sec. 3, we want the constraints expressed in terms of state variables and differential equations. Let $\underline{s}(\underline{w}(t), u(t), t)$ be the vector composed of the $s_k(\underline{w}(t), u(t), t)$ of Eq. (5.5). Define the state vector, $\underline{r}(t)$ by

$$\frac{d}{dt}\underline{r}(t) = \underline{s}(\underline{w}(t), u(t), t) \tag{5.6}$$

The constraints of Eq. (5.5) then become constraints on $\underline{r}(T)$. In Sec. 4 we saw that constraints on the state variables, $I(t)$ at time T may arise through the choice of criteria. Here we see that similar constraints on the state variables $\underline{r}(t)$ at the T result from the physical restrictions imposed on the transmitted signal.

Another way to handle constraints is to use a criteria which combines $I(T)$ with some property of the signal; for example, maximize

$$\ell(T) = \log|I(T)| - \int_0^T u^2(t)dt .$$

However, such an approach does not appear natural in most problems of interest.

6. PONTRYAGIN'S MAXIMUM PRINCIPLE

Pontryagin's Maximum Principle is a concise statement of much of the classical theory of the calculus of variations. However, it also encompasses a wider range of problems such as those with peak amplitude constraints. The Maximum Principle is discussed with neither proof nor rigor and only to a rather limited degree of generality.

Let $\underline{q}(t)$ denote a state vector with elements $q_j(t)$, $j = 1, \ldots, \rho$, defined by the following system of first-order differential equations.

$$\frac{dq_j(t)}{dt} = g_j(\underline{q}(t), u(t), t) \qquad j = 1, \ldots, \rho \tag{6.1}$$

or in complete vector notation

$$\frac{d\underline{q}(t)}{dt} = \underline{g}(\underline{q}(t), u(t), t) \tag{6.2}$$

when the g_j are the elements of $\underline{g}$. In Section 3, 4, and 5 we introduced the state variables which define the information $\mathcal{I}(t)$; the criterion, $\mathcal{L}(t)$, and the constraints, $\underline{r}(t)$ and $\underline{w}(t)$. These state variables are related to the waveform $x(t)$ and the control function $u(t)$. One element of $\underline{q}(t)$ is to be associated with each element of $\mathcal{I}(t)$, $\mathcal{L}(t)$, $\underline{r}(t)$ and $\underline{w}(t)$. The differential equation, Eq. (6.2), is thus composed of equations such as Eqs. (3.17), (4.3), (5.3) and (5.6). The $A(t)$, $B(t)$, and $C(t)$ of Eq. (3.17) are to be expressed, where necessary, in terms of the $\underline{w}(t)$ and control $u(t)$. For example, Eq. (3.18) is a special case which illustrates the dependence on the signal $x(t)$ which is, in turn, to be defined by the $\underline{w}(t)$ and $u(t)$.

Define the vector $\underline{p}(t)$ with elements $p_j(t)$, $j = 1, \ldots, \rho$, by the equations,

$$H = \underline{p}'(t)\underline{g}(\underline{q}(t), u(t), t) \tag{6.3}$$

where the scalar quantity H is the Hamiltonian* and

$$\frac{d}{dt} p_j(t) = - \frac{\partial H}{\partial q_j(t)} \qquad j = 1, \ldots, \rho$$

or in vector notation

$$\frac{d\underline{p}(t)}{dt} = - \frac{\partial H}{\partial \underline{q}(t)} \tag{6.4}$$

where the right hand side is the vector whose elements are the partials of H with respect to the $q_j(t)$. The $p_j(t)$ are called here the adjoint variables although the terminology costate variables or Lagrange multipliers is also employed in the literature. Substituting Eq. (6.3) into Eq. (6.4) gives,

$$\frac{d\underline{p}(t)}{dt} = - G_{\underline{q}}(\underline{q}(t), u(t), t)\underline{p}(t) \tag{6.5}$$

where $G_{\underline{q}}$ is the $\rho \times \rho$ matrix whose elements are

$$\frac{\partial g_k(\underline{q}(t), u(t), t)}{\partial q_j(t)} .$$

Equation (6.5) is the adjoint equation for the linearized version of Eq. (6.2) and hence the name, adjoint variables, for $\underline{p}(t)$. Equation (6.2) can be rewritten as

$$\frac{d\underline{q}(t)}{dt} = \frac{\partial H}{\partial \underline{p}(t)} .$$

Note that $I(t)$ is a symmetric matrix while $\underline{q}(t)$ is a vector. It would be possible to write the significant elements of $I(t)$ in a column vector but it has proven

* This Hamiltonian is closely related to the Hamiltonian of classical physics where the p'_js are the momentum variables.

conceptually valuable to maintain $\mathcal{I}(t)$ as a separate entity; namely, a matrix. Since this is rather unconventional, we indicate the form of the resulting equations and how they are manipulated when $\mathcal{I}(t)$ is given by Eq. (3.17). Define $\underline{q}_1(t)$ as the vector containing all the state variables except those in $\mathcal{I}(t)$. Then

$$\frac{d\underline{q}_1(t)}{dt} = \underline{g}_1(\underline{q}_1(t), \mathcal{I}(t), u(t), t) \ .$$

Let $\underline{p}_{\underline{q}_1}(t)$ be the adjoint variables associated with $\underline{q}_1(t)$. Since $\mathcal{I}(t)$ is a symmetric matrix, we represent its corresponding adjoint variables by the symmetric matrix, $P_{\mathcal{I}}(t)$. The Hamiltonian of Eq. (6.3) is now written in the form*

$$H = H_1 + H_2$$

where H_1 is the conventional form of Eq. (6.3)

$$H_1 = \underline{p}'_{\underline{q}_1}(t)\underline{g}_1(\underline{q}_1(t), \mathcal{I}(t), u(t), t)$$

while H_2 is by inspection of the form

$$H_2 = \mathrm{tr}\ P_{\mathcal{I}}(t)\ \frac{d\mathcal{I}(t)}{dt}$$

where $\frac{d\mathcal{I}(t)}{dt}$ is Eq. (3.17) written as a function of $\underline{q}_1(t), \mathcal{I}(t), u(t)$ and t. Stating Eq. (6.4) in matrix notation,

$$\frac{dP_{\mathcal{I}}(t)}{dt} = - \frac{\partial H}{\partial \mathcal{I}(t)}$$

where the right hand side denotes the matrix of partials of H with respect to the elements of $\mathcal{I}(t)$. Then by performing the necessary partial differentiations, it can be

* These H's obviously bear no relation to the hypotheses H_1 and H_2.

shown without too much difficulty that

$$\frac{dP_{\mathcal{I}}(t)}{dt} = [A(t)+B(t)\Sigma_{\underline{\xi}_1} B'(t)\mathcal{I}(t)]\, P_{\mathcal{I}}(t) + P_{\mathcal{I}}(t)\,[A'(t)+\mathcal{I}(t)B(t)\Sigma_{\underline{\xi}_1} B'(t)] - \frac{\partial H_1}{\partial \mathcal{I}(t)} \,. \tag{6.6}$$

The last term of Eq. (6.6) and the equations for $\frac{dp_{\underline{q}_1}(t)}{dt}$ depend on the particular problem. It should be emphasized that there is no new concept involved in this combination of vector and matrix differential equations. It is simply a convenient notation for dealing with a particular system of ρ equations.

As discussed in Sec. 5, a strict inequality constraint

$$\epsilon_1 \le u(t) \le \epsilon_2 \tag{6.7}$$

on the control function may exist. Let $u^o(t)$ denote the control function which maximizes $q_1(T)$ subject if necessary to Eq. (6.7) where we associate the criterion state variable, $\ell(t)$, with $q_1(t)$.

Now consider Pontryagin's Maximum Principle. The Hamiltonian of Eq. (6.3) is a function of $t, \underline{p}(t), \underline{q}(t)$ and $u(t)$; that is,

$$H = H(\underline{p}(t), \underline{q}(t), u(t), t) \,. \tag{6.8}$$

The Maximum Principle tells us that $u^o(t)$ must minimize H, subject if necessary to the constraint of Eq. (6.7). This means that u^o, expressed as a function of $\underline{p}(t), \underline{q}(t)$, and t, must result in the minimum H with respect to variations in u. Section 7 will illustrate this minimization when a constraint such as Eq. (6.7) is necessary. When Eq. (6.7) is not required, the necessary conditions on $u^o(t)$ are obtained from $\partial H/\partial u(t)$. This is illustrated in Sec. 8. Section 8 actually employs a vector control function, $\underline{u}(t)$, but this is a direct extension; that is, $\underline{u}^o(t)$ must absolutely minimize H.

The preceding is not an actual statement of the Maximum Principle but rather a weaker corollary drawn from it. This corollary does not always provide all the desired information. For example, $\underline{p}(t)$ and $\underline{q}(t)$ may be such that H is independent

over some finite period of time of the value of u or its sign or its magnitude. This is called the singular case and can arise in certain problems of waveform design. However, for many problems of interest, the control which minimizes the Hamiltonian is unique and can be completely expressed in terms of $\underline{p}(t), \underline{q}(t)$ and t. Thus u(t) can be eliminated from Eqs. (6.2) and (6.4) to give a system of 2ρ first-order differential equations in the 2ρ variables, $\underline{p}(t)$ and $\underline{q}(t)$. Solution for the resulting $\underline{p}(t)$ and $\underline{q}(t)$ gives $u^o(t)$ and thus the optimum waveform.

Now consider the important question of boundary conditions for Eq. (6.2) and Eq. (6.4). We need 2ρ boundary conditions to specify a solution. $\underline{q}(t)$ is composed of $\mathcal{I}(t), \underline{w}(t), \underline{r}(t)$ and $\ell(t)$. $\mathcal{I}(0)$ is the a priori information, while $\underline{w}(0), \underline{r}(0)$, and $\ell(0)$ are part of the model. Thus, $\underline{q}(0)$ provides ρ boundary conditions. The other ρ conditions are obtained at time T. Some elements of $\underline{q}(T)$ are usually specified or functionally related by the criterion and constraints. However, certain of the elements of q(T) are usually unspecified. In the radar problem, a possible criterion is the minimization of $|\mathcal{I}(T)|$, with no specification on the elements of $\mathcal{I}(T)$. Thus, $\underline{q}(T)$ may furnish some but not all of the remaining ρ boundary conditions. The rest are obtained from the transversality conditions. Although these transversality conditions can become complex, the underlying principle is simple. Suppose the problem specifies the values of ρ_1 elements of $\underline{q}(T)$. The remaining $\rho-\rho_1$ elements are simply chosen such that the criterion state variable, $q_1(T)$ (or $\ell(T)$), is maximum. These are the transversality conditions. In more general problems where ρ_1 equations or inequalities on q(T) are specified, the transversality conditions simply become more complicated. The case of a free T is obviously a special case.

Thus for many problems, the Maximum Principle defines the maximum information signal in terms of the solution of a system of first-order differential equations with split boundary conditions; i.e., a two-point boundary value problem. In some cases the equations can be integrated in closed form to give a system of algebraic or transcendental equations. If $\underline{p}(0)$ were known, integration on a computer would be simple. Thus iterative search techniques can be employed on $\underline{p}(0)$ to determine which provides the desired relationship among the elements of q(T). "Insight" may enable changes of variables which simplify the integration or the beautiful Hamilton-Jacobi

theory of classical physics can be employed. This leads to the theory of Poisson and Lagrange brackets and canonical transformations. The Hamilton-Jacobi theory is closely related to Dynamic Programming which is still another way to approach the basic problem. Thus we are faced with a host of possible methods for obtaining explicit solutions. Unfortunately the best choice depends on both the problem and the computing facilities available. Each new situation must usually be treated as an entity unto itself.

The Maximum Principle can be extended to include phase coordinate constraints such as discussed in Sec. 5. However, this extension constitutes an order of magnitude increase in difficulty when it comes to actual solutions. It can also be extended in discrete time systems.

The above discussions are obviously not complete and the interested reader must resort to either previous knowledge or the literature if he wishes to actually solve problems. Two references on the Maximum Principle employed by the author are Refs. 20 and 21. Reference 22 contains an extensive bibliography. The reader who, like the author, is more interested in understanding the use rather than the proof of the theory, may find the classical calculus of variations as it was developed for physics very valuable. Reference 23 is a good general reference while Ref. 24 is a well-written work which relates optimal control theory to these classical approaches.

7. THE RADAR EXAMPLE

Consider the radar problem first stated in Sec. 2. The waveform is to be constrained with respect to both total energy and peak amplitude. Thus, following Sec. 5, we set

$$x(t) = u(t) \tag{7.1}$$

$$\frac{dr(t)}{dt} = u^2(t) \tag{7.2}$$

$$0 \leq u(t) \leq \epsilon_1 \qquad \text{(peak amplitude constraint)} \tag{7.3}$$

$$r(T) = \epsilon_2 \qquad \text{(energy constraint)} \ . \tag{7.4}$$

Since $\mathcal{I}(t)$ depends only on $x^2(t)$, we have assumed in Eq. (7.3) that $x(t)$ is always positive. Substituting Eq. (7.1) into Eq. (3.18) gives

$$\frac{d}{dt}\mathcal{I}(t) = \frac{u^2(t)}{\sigma^2}\,\underline{\phi}(t)\underline{\phi}'(t) \tag{7.5}$$

where

$$\underline{\phi}(t) = \begin{bmatrix} 1 \\ \\ t \end{bmatrix}.$$

For the sake of illustration, assume the criterion is the maximization of the determinant of $\mathcal{I}(T)$. Thus, following Sec. 4

$$\ell(T) = \log \left|\mathcal{I}(T)\right| \tag{7.6}$$

and using Eq. (7.5) and Eq. (4.4)

$$\frac{d\ell(t)}{dt} = \frac{u^2(t)}{\sigma^2}\,\mathrm{tr}\ \mathcal{I}^{-1}(t)\,\underline{\phi}(t)\underline{\phi}'(t) \quad . \tag{7.7}$$

In addition to the boundary conditions of Eqs. (7.4) and (7.6) we impose the conditions,

$$\mathcal{I}(0) = 0$$

$$r(0) = 0$$

$$\ell(0) = 0 \quad .$$

The differential equation for $\underline{q}(t)$, Eq. (6.2), is the combination of Eqs. (7.2), (7.5) and (7.7). Let $P_{\mathcal{I}}(t)$ be the adjoint (matrix) corresponding to $\mathcal{I}(t)$ while $p_r(t)$ and $p_\ell(t)$ correspond to $r(t)$ and $\ell(t)$ respectively. The Hamiltonian of Eq. (6.3) then becomes by inspection,

$$H = \frac{u^2(t)}{\sigma^2} \operatorname{tr} P_{\mathcal{I}}(t)\underline{\phi}(t)\underline{\phi}'(t) + \frac{u^2(t)}{\sigma^2} p_\ell(t) \operatorname{tr} \mathcal{I}^{-1}(t)\underline{\phi}(t)\underline{\phi}'(t) + p_r(t)\, u^2(t) \quad . \qquad (7.8)$$

Now, using Eq. (6.4),

$$\frac{dp_\ell(t)}{dt} = - \frac{\partial H}{\partial \ell(t)} = 0$$

and

$$\frac{dp_r(t)}{dt} = - \frac{\partial H}{\partial r(t)} = 0 \ .$$

Thus $p_\ell(t)$ and $p_r(t)$ are constants, p_ℓ and p_r, respectively. Repeated use of the matrix equations such as

$$d\mathcal{I}^{-1} = -\mathcal{I}^{-1} d\mathcal{I}\,\mathcal{I}^{-1}$$

$$\operatorname{tr} P\, \underline{\phi}\,\underline{\phi}' = \underline{\phi}' P\, \underline{\phi}$$

gives

$$\frac{dP_{\mathcal{I}}(t)}{dt} = - \frac{\partial H}{\partial \mathcal{I}(t)}$$

$$= - \frac{u^2(t)}{\sigma^2} p_\ell \frac{\partial}{\partial \mathcal{I}(t)} \underline{\phi}'(t) \mathcal{I}^{-1}(t) \underline{\phi}(t)$$

$$= \frac{u^2(t)}{\sigma^2} p_\ell \mathcal{I}^{-1}(t) \underline{\phi}(t) \underline{\phi}'(t) \mathcal{I}^{-1}(t)$$

$$= - p_\ell \frac{d}{dt} \mathcal{I}^{-1}(t) .$$

Thus,

$$P_{\mathcal{I}}(t) + p_\ell \mathcal{I}^{-1}(t) = D \qquad \text{a constant matrix}$$

and the Hamiltonian of Eq. (7.8) becomes

$$H = u^2(t) \zeta(t)$$

where

$$\zeta(t) = \frac{1}{\sigma^2} \underline{\phi}'(t) D \underline{\phi}(t) + p_r .$$

From the Maximum Principle we know that the optimum control, $u^o(t)$, must absolutely minimize H subject to the constraint of Eq. (7.3). Thus,

$$u^o(t) = \begin{cases} \epsilon_1 & \zeta(t) < 0 \\ 0 & \zeta(t) > 0 . \end{cases}$$

Now $\zeta(t)$ is of the form

$$\zeta(t) = \zeta_0 + \zeta_1 t + \zeta_2 t^2 .$$

Thus $\zeta(t)$ can change sign at most twice which means $u^o(t)$ and thus the optimum waveform

x(t) is "on-off" with at most two periods of full transmission. *

The actual $u^o(t)$ could be found from the boundary and transversality conditions and a solution of the two-point boundary value problem. However, it is deemed easier to investigate all "on-off" signals which satisfy the constraints and which have at most two periods of transmission to see which provides the maximum $|I(T)|$. This can be considered a change of variables. The results of this investigation will be reported in a later report which also considers other criteria and the case of an accelerating target. The actual results were obtained from a digital computer but because of the small number and limited range of the parameters which characterize the possible optimum waveforms, very little computation time is involved. These results also show that the choice of criteria usually affects only the switch times, not the form of the optimum.

Since our optimum is "on-off", it is an infinite bandwidth waveform and can never be exactly implemented. However, if the rise time of our band-limited transmitter is small compared to the total time duration T, the mathematically optimum waveform provides a good basis for design.

It is important to remember that the radar model of Sec. 2 is limited in scope and was chosen primarily to illustrate the technique of optimum waveform design. Therefore the above conclusions on the shape of optimum radar waveforms are only examples of the type of conclusions possible and are actually applicable to only special types of real radar problems.

* This is a valid conclusion only after verification that $\xi(t)$ is not identically zero for all time.

8. THE COMMUNICATION EXAMPLE

Consider the communication problem first stated in Sec. 2. The differential equation for $\mathcal{I}(t)$ is obtained from Eq. (3.27) as

$$\frac{d\mathcal{I}(t)}{dt} = \left(\frac{1}{\sigma^2}\right)\underline{x}(t)\underline{x}'(t) \tag{8.1}$$

where $\underline{x}(t)$ is the vector of the waveforms $x_k(t)$. Each of the M possible waveforms are to be constrained with respect to total energy and bandwidth. To handle the energy constraints, we want to constrain

$$\int_0^T x_k^2(t)dt \qquad k = 1, \ldots, M \ .$$

However, as discussed in Sec. 4, the main diagonal elements of $\mathcal{I}(T)$ are proportional to the total waveform energy for the case of white observation noise. Thus, the energy constraints are handled by

$$\iota_{kk}(T) \le \epsilon_1 \ . \tag{8.2}$$

As discussed in Sec. 5, there are various ways to handle a bandwidth constraint and the choice depends on the particular problem. For this example, we constrain the second moment of the energy spectrum. Thus we define*

$$\frac{d\underline{x}(t)}{dt} = \underline{u}(t) \tag{8.3}$$

$$\frac{d\underline{r}(t)}{dt} = \underline{u}^2(t) \tag{8.4}$$

where $\underline{u}(t)$ is the vector of the M controls, $u_k(t)$, one for each $x_k(t)$ and $\underline{u}^2(t)$ is (unconventional notation) the vector of the $u_k^2(t)$. The bandwidth constraint is then

* Section 5 employed an auxiliary state variable $\underline{w}(t)$. However, here it is simpler to work with $\underline{x}$ directly and consider $\underline{x}$ a state vector.

$$r_k(T) \le \epsilon_2 \qquad k = 1, \ldots, M \tag{8.5}$$

where the $r_k(t)$ are the elements of $\underline{r}(t)$.

For the sake of illustration, we use Eqs. (4.1), (4.5) and (4.6). Equation (4.1) is *

$$\iota_{kk}(T) \ge \epsilon_3 \; . \tag{8.6}$$

Rewriting Eq. (4.5) and Eq. (4.6)

$$\iota_{kk}(T) - 2\iota_{jk}(T) + \iota_{jj}(T) \ge \ell(T) \tag{8.7}$$

$$\frac{d\ell(t)}{dt} = 0 \; . \tag{8.8}$$

In addition to the boundary conditions of Eqs. (8.5), (8.6) and (8.7) we impose,

$$\underline{r}(0) = 0$$

$$I(0) = 0$$

$$\underline{x}(0) = 0$$

$$\underline{x}(T) = 0 \; .$$

The differential equation for $\underline{q}(t)$, Eq. (6.2), is the combination of Eqs. (8.1), (8.3), (8.4) and (8.8). Let $P_I(t)$ denote the adjoint (matrix) corresponding to $I(t)$, $\underline{p}_r(t)$ the adjoint (vector) corresponding to $\underline{r}(t)$, $\underline{p}_x(t)$ the adjoint (vector) corresponding to $\underline{x}(t)$ and $p_\ell(t)$ the adjoint (scalar) corresponding to $\ell(t)$. The Hamiltonian then becomes

$$H = \frac{1}{\sigma^2} \operatorname{tr} P_I(t)\underline{x}(t)\underline{x}'(t) + \underline{p}_x'(t)\,u(t) + \underline{p}_r'(t)\,\underline{u}^2(t) + (0)\,p_\ell(t).$$

* Obviously ϵ_3 of Eq. (8.6) cannot be greater than ϵ_1 of Eq. (8.2).

Now

$$\frac{d\underline{p}_r(t)}{dt} = -\frac{\partial H}{\partial \underline{r}(t)} = 0$$

$$\frac{dP_I(t)}{dt} = -\frac{\partial H}{\partial I(t)} = 0$$

which means $\underline{p}_r(t)$ and $P_I(t)$ are constants, $\underline{p}_r$ and P_I respectively. Similarly

$$\frac{d\underline{p}_x(t)}{dt} = -\frac{\partial H}{\partial \underline{x}(t)} = -\frac{2}{\sigma^2} P_I \underline{x}(t) \tag{8.9}$$

Let $P_{\underline{r}}^{(-1)}$ denote (unconventional notation) the diagonal matrix formed by the reciprocals of the elements of $\underline{p}_r$.

Minimization of the Hamiltonian with respect to $\underline{u}(t)$ is obtained by partial differentiation. This gives for the optimum,

$$\underline{u}^o(t) = -\frac{1}{2} P_{\underline{r}}^{(-1)} \underline{p}_x(t) . \tag{8.10}$$

Substituting Eq. (8.10) into Eq. (8.3) and combining with Eq. (8.9) gives

$$\frac{d}{dt}\begin{bmatrix} \underline{x}(t) \\ \\ \underline{p}_x(t) \end{bmatrix} = \begin{bmatrix} O \\ \\ \frac{-2}{\sigma^2} P_I \end{bmatrix} \begin{bmatrix} -\frac{1}{2} P_{\underline{r}}^{(-1)} \\ \\ O \end{bmatrix} \begin{bmatrix} \underline{x}(t) \\ \\ \underline{p}_x(t) \end{bmatrix} \tag{8.11}$$

or

$$\underline{x}(t) = -\frac{1}{2} P_{\underline{r}}^{(-1)} \int_0^t \underline{p}_x(\tau) d\tau \tag{8.12}$$

$$\frac{d^2}{dt^2} \underline{p}_x(t) = \frac{1}{\sigma^2} P_I P_{\underline{r}}^{(-1)} \underline{p}_x(t) \tag{8.13}$$

Equation (8.11) or Eqs. (8.12) and (8.13) specify the form of the optimum waveforms. $P_{\underline{r}}^{(-1)}$ and P_I are constant. Thus the optimum waveforms are generated by a system of 2M, first-order, constant coefficient, linear, time invariant, differential equations (Eq. (8.11)) or by a simple system of M second-order, constant coefficient, linear, time invariant differential equations (8.13) followed by M integrators (Eq. (8.12)). Such systems and their corresponding matched filter receivers are easy to build and thus the optimum results in a system which can be readily implemented.

There remains, of course, the determination of the initial conditions, $\underline{p}_{\underline{x}}(0)$, and the values of P_I and $\underline{p}_{\underline{r}}$. These quantities are specified by the boundary conditions and transversality conditions and of course, depend on Eqs.(8.1) and (8.4). Since all the necessary equations can be integrated in closed form, the two-point boundary value problem can be circumvented if desired. However, we have not attempted an actual solution.

As with the radar example, remember that the basic model which leads to the preceding conclusions was chosen to illustrate the theory, not for realism.

9. DISCUSSION

We have relied primarily on simple examples to indicate the breadth of the methodology of optimum signal design. Therefore, let us briefly discuss the complete range of applications. Obviously only samplings of possible criterion and constraints were given. Discrete time problems can be solved. The complete properties of the received signal are determined by its probability distribution and a finite number of state variables (such as an information matrix) can completely determine this distribution only in special cases. Thus it will sometimes be necessary to directly control the probability distributions of the received signal and this can lead to partial differential equations and infinite dimensional state systems. The necessary optimization theory is still in its infancy but work is progressing. We considered only the design of deterministic waveforms.* However, the basic technique of combining a state variable model with Pontryagin's Maximum Principle is equally applicable to the design of a probability distribution, $f(x)$. The independent variable, t, in $x(t)$, is simply replaced by x, in $f(x)$. The design of optimum power spectra as discussed in Ref. 1 can also be done in a similar fashion. Note that the maximum principle is essential for such problems as the classical calculus of variations does not allow the incorporation of the necessary constraint that the probability density (or the power spectra) never be negative.

Our general discussions ended with the optimum waveform specified as a solution to a two-point boundary value problem; that is, the optimum waveform was parameterized by a finite number of parameters (the boundary conditions) which still remained to be evaluated. It is interesting to compare this result with the design technique wherein the range of possible waveforms is initially parameterized and then the optimum values of the parameters determined. (For example, the waveform is restricted to be a linear combination of some chosen, finite, set of orthonormal functions.) Our parameterization is automatically specified by the inherent dynamics of the information, the chosen criterion and the applied constraints. Since the resulting parameters are natural to the problem, solution for their optimum values is easier as

* These include the case of some stochastic signals as illustrated by the laser radar example of Sec. 3.

the theory discussed at the end of Sec. 6 is available. In addition, this parameterization of the optimum signal has its own intrinsic value. For example, we learn immediately that a pulse radar is better than a CW radar and that the optimum communication waveforms are easy to generate. *

We have drawn heavily on the concepts and results of optimal control theory. Although analogies abound, there is one important and felicitous difference. In control problems, the optimum control is usually wanted as a function of the state of the system; that is, a feedback control law. In signal design we are satisfied with the waveform (control) as a function of time. Thus, many techniques (such as iterative search on boundary conditions) which are inappropriate for most realistic control problems can be used for waveform design.

The basic principles of optimum waveform design are very simple. All work is done in the time domain. The pertinent properties of the received signal are explicitly displayed by the state variable model of the information. Signal constraints and the chosen criteria are handled in a natural manner and automatically incorporated and displayed through the appropriate state variables. The resulting system of differential (difference) equations succumbs easily to the calculus of variations (Pontryagin's Maximum Principle) which provides necessary conditions the optimum waveforms must satisfy. Actual evaluation of the optimum waveform may not be easy but the equations are well adapted to modern computers and many avenues are open. Realistic problems can be solved.

* Naturally, these statements apply only in the context of the limited models used here!

APPENDIX

The definition of information used in Sec. 3 is not the definition commonly employed in communication theory. Therefore, we shall relate the two concepts. Much of the discussion is extracted from the first two chapters of Ref. 6, often nearly verbatim.

Assume the random variable, z, is observed and that a decision between the two hypotheses,

$$\begin{aligned} H_1 &: f_1(z) \\ H_2 &: f_2(z) \end{aligned} \tag{A.1}$$

must be made. It is well-known in communication and statistical theory, (see Refs. 6, 7, 17) that the likelihood ratio, $f_1(z)/f_2(z)$, forms the basis for most hypothesis tests and is also fundamental to the theory of parameter estimation. For a particular observation, z, the larger $f_1(z)/f_2(z)$ the more likely we are to decide in favor of H_1. Thus, $\log f_1(z)/f_2(z)$ measures, in some sense, the information gained in favor of H_1 over H_2. To show why $\log f_1(z)/f_2(z)$ is an especially appropriate measure, assume the existence of a priori probabilities, $P(H_j)$, $j = 1, 2$, on the two hypotheses. Let $P(H_j/z)$ $j = 1, 2$, denote the conditional probability of H_j given the sample z. By Bayes' theorem,

$$P(H_j/z) = \frac{P(H_j)\, f_j(z)}{P(H_1)\, f_1(z) + P(H_2)\, f_2(z)} \qquad j = 1, 2$$

from which

$$\log\frac{f_1(z)}{f_2(z)} = \log\frac{P(H_1/z)}{P(H_2/z)} - \log\frac{P(H_1)}{P(H_2)} \quad .$$

Thus, $\log f_1(z)/f_2(z)$ is the difference between the logarithm of the "odds" in favor of H_1 over H_2, as calculated before and after observing z. The question of when and where a priori distributions such as $P(H_j)$ should be used is the subject of many debates but fortunately it need not be pursued here. $\log f_1(z)/f_2(z)$ is well-defined, whether or

whether not the $P(H_j)$ exist. If the $P(H_j)$ are known, their effect is additive.

The mean information, $I(1:2/z)$ as defined in Sec. 3 is simply the average value of $\log f_1(z)/f_2(z)$ assuming H_1 is true.

$$I(1:2/z) = \int f_1(z) \log \frac{f_1(z)}{f_2(z)} \, dz \; . \tag{A.2}$$

Now, in general

$$I(1:2/z) \neq I(2:1/z) \; .$$

This sometimes makes the concept of divergence, $J(1:2/z)$, a useful extension where

$$J(1:2/z) = I(1:2/z) + I(2:1/z)$$

$$= \int (f_1(z) - f_2(z)) \log \frac{f_1(z)}{f_2(z)} \, dz \; .$$

The divergence, $J(1:2/z)$ has all the properties of a distance (or metric) except for the triangle irregularity. In this context, $I(1:2/z)$ can be considered a diverted divergence. The choice between using $J(1:2/z)$ and $I(1:2/z)$ depends on the problem. *

Now consider the definition of information commonly employed in communication theory (see Sec. 3.5 of Ref. 25). Assume x is transmitted with probability density $g(x)$ and z is received with probability density $h(z)$. Let $f(x, z)$ be the joint distribution. The quantity, I, given by

$$I = \int \int f(x, z) \log \frac{f(x, z)}{g(x)h(z)} \, dx \, dz \tag{A.3}$$

is called the mean information in z about x, or the mean mutual information between x and z. The definition of Eq. (A.3) can be related to the rate of transmission of information and the channel capacity, (see Ref. 25). Section 2.5 of Ref. 6 also discusses

---

* In the communication and radar problems of Sec. 2, $I(1:2/z) = I(2:1/z)$ and thus no decision was required.

these concepts.

The mutual information of Eq. (A.3) is a special case of Eq. (A.2) in the following sense. If

H_1: x and z dependent with distribution f(x, z)

H_2: x and z independent with distributions g(x) and h(z).

Then,

$$I = I(1:2/z).$$

In Sec. 3 we developed the information matrix concept using the definition of Eq. (A.2) and in Sec. 4 discussed the choice of a scalar function of this matrix which is to be maximized. Since Eq. (A.3) is a scalar measure of information, it can be considered simply another criterion which could have been included in Sec. 4.

Equation (A.3) could be employed in the communication example of Sec. 2 if we assign an a priori probability distribution to the α_k. However, such an approach obscures the basic nature of the problem as the various trade-offs between type 1 and 2 errors are not displayed. Thus, specializations of Eq. (A.2) other than Eq. (A.3) were deemed more appropriate.

REFERENCES

1. R.M. Lerner, "Representation of Signals," "Design of Signals," Chapters 10 and 11 of Lectures on Communication System Theory, by E.J. Baghdady (McGraw-Hill Book Co., 1961).

2. R.W. Lucky and J.C. Hancock, "On the Optimum Performance of N-ary Systems Having Two Degrees of Freedom," IRE Transactions on Communication Systems, CS-10, pp. 185-192 (June 1962).

3. G. Luchs, "Optimization of Signal Waveforms," IEEE Transactions on Information Theory (April 1963).

4. E.J. Kelly, "The Radar Measurement of Range, Velocity and Acceleration," IRE Transactions on Military Electronics (April 1961).

5. E.J. Kelly, I.S. Reed, and W.L. Root, "The Detection of Radar Echoes in Noise," 8, Journal of the Society of Industrial and Applied Mathematics (June 1960).

6. S. Kullback, Information Theory and Statistics (John Wiley and Sons, New York, N.Y., 1959).

7. S.S. Wilks, Mathematical Statistics (John Wiley and Sons, New York, N.Y., 1962).

8. L.A. Zadeh, C.A. Desoer, Linear System Theory: The State Space Approach (McGraw-Hill Book Co., New York, N.Y., 1963).

9. H.W. Bode and C.E. Shannon, "A Simplified Derivation of Linear Square Smoothing and Prediction Theory," Proceedings of IRE, 38, pp. 417-425 (April 1950).

10. R.E. Kalman and R.S. Bucy, "New Results in Linear Filtering and Prediction Theory," Journal of Basic Engineering ASME Transactions, 83 D (1961).

11. R.E. Kalman, "New Methods and Results in Linear Prediction and Filtering Theory," TR-61-1, RIAS, Baltimore, Maryland (November 1960).

12. F.C. Schweppe, "An Introduction to Estimation Theory for Time Varying Systems," 22G-15(U), MIT, Lincoln Laboratory (September 1963).

13. D.J. Sakrison, "Comparison of Linear and Nonlinear Estimation Methods for Signals Containing Unknown Parameters," JA-2251(U), MIT, Lincoln Laboratory (October 1963).

14. L. Jones, F. Schweppe, "A Random Acceleration Model for Filtering Polynomial-Like Signals," 28G-3(U), MIT, Lincoln Laboratory (25 October 1963).

15. B. Reiffen, H. Sherman, "An Optimum Demodulator for Poisson Processes: Photon Source Detectors," Proceedings of IEEE, 51, No. 10 (October 1963), also AP-25(U) and 25G-16(U), MIT, Lincoln Laboratory (20 February 1963).

16. R. Price, "Optimum Detection of Random Signals in Noise, with Application to Scatter-Multipath Communication," Trans. I. R. E., IT-2, No. 4, 125 (1956).

17. C.W. Helstron, Statistical Theory of Signal Detection (Pergamon Press, New York, 1960).

18. L. A. Zadeh, "Optimality and Non-Scalar-Valued Performance Criteria," AC-8, IEEE Transactions on Automatic Control (January 1963).

19. C. L. Cuccin, Harmonics, Sidebands and Transients in Communication Engineering (McGraw-Hill Book Co., 1952).

20. L.S. Pontryagin, V.G. Boltyanskii, R.V. Gamkrelidze, E.F. Mishchenko, The Mathematical Theory of Optimal Processes (Interscience Publishers, New York, 1962).

21. L.I. Rozonoer, "The L.S. Pontryagin Maximum Principle in the Theory of Optimal Systems, I, II, and III, Avtomatika i Telemekh., 20, 1320-1334, 1441-1458, 1561-1578 (1959). English translation in Automation and Remote Control, 20, 1288-1302, 1405-1421, 1517-1532 (1959).

22. M. Athanassiades, "Optimal Control for Linear Time-Invariant Plants with Time, Fuel, and Energy Constraints," IEEE Transactions on Applications and Industry, 81, pp. 321-325 (January 1963), also JA-1923(U), MIT, Lincoln Laboratory (January 1963).

23. C. Lanczos, The Variational Principles of Mechanics (University of Toronto Press, Toronto, Canada, 1949).

24. R. E. Mortensen, "A Synopsis of Optimal Control Theory," Series No. 60, Issue No. 487, Electronics Research Laboratory, University of California, Berkeley, California (November 14, 1962).

25. P. M. Woodward, Probability and Information Theory, with Applications to Radar (McGraw-Hill Book Co., New York, 1953).

DISTRIBUTION LIST

Director's Office

W. B. Davenport, Jr.

Division 2

C. R. Wieser
S. H. Dodd
F. C. Frick

Group 22

V. A. Nedzel
W. I. Wells

R. J. Bergemann
L. F. Cianciolo
D. F. Clapp
A. Freed
L. A. Globus
A. A. Grometstein
J. H. Hartogensis
R. C. Holland
C. Hopkins
M. M. Klein
E. Korngold
W. Z. Lemnios
L. M. Meixsell
L. J. Murray
B. Nanni
C. E. Nielsen
J. H. Pannell
E. W. Pike
D. Reiner
J. Rheinstein
J. Salerno
B. J. Schafer
A. F. Smith
J. S. Strano
H. Sussman
J. L. Vernon

Group 28

J. A. Arnow
J. F. Nolan

A. W. Armenti
C. R. Arnold
M. Athans (5)
T. C. Bartee
R. N. Davis
P. L. Falb
L. A. Gardner, Jr.
P. L. Grant
D. A. Koniver
H. K. Knudsen
J. B. Lewis
H. E. Meily
A. J. Morency
H. C. Peterson
F. C. Schweppe (100)
J. Slesers
J. M. Winett
P. E. Wood

Group 41

D. L. Clark
E. M. Hofstetter
E. J. Kelly, Jr.
J. B. Resnick
H. M. Jones
R. E. Vanderschuur

Group 42

W. W. Camp
A. A. Galvin
J. Margolin

Group 44

J. L. Allen

Group 45

J. P. Perry
W. W. Ward

Group 62

L. I. Bluestein
P. R. Drouilhet
K. Jordan
R. S. Kennedy
I. L. Lebow

Group 63

R. Lerner
J. Max

Group 64

P. E. Green
J. H. Holsinger
M. J. Levin
R. Price

Group 65

H. D. Sherman
B. Reiffen
H. L. Vantrees

Group 66

R. T. Prosser

MIT - Dept. of Elec. Engineering

R. W. Brockett
Room 35-006

G. C. Newton
Room 32-106

J. M. Wozencraft
Room 26-347

G. D. Zames
Room 32-102

Hq. ESD (Maj. S. Jeffery ESRC)
L. G. Hanscom Field
Bedford, Mass.

Hq. ESD (Lt. J. Egan)
L. G. Hanscom Field
Bedford, Mass.

FCS: sm
16 January 1964